Statistique descriptive
Chapitre 6 : Tableaux statistiques à doubles entrées
Mesures d'Indépendance et de corrélation

Mohamed BOUINCHA
m.bouincha@gmail.com
Institut National de Statistique et d'Économie Appliquée

Année universitaire : 2024/2025

INSEA

Sommaire

1. Types de croisements entre caractères relatifs aux même unités
2. Tableaux statistiques à doubles entrées
3. Distributions statistiques relatives au tableau à doule entrées
4. Notion d'Indépendance, de Corrélation et de liaison fonctionnelle
5. Indépendance et corrélation pour les variables quantitatives
6. Liaison entre deux variables qualitatives
7. Relation entre une variable qualitative et une variable quantitative

INSEA

Types de croisements entre caractères relatifs aux même unités

- Exemple de base de données statistique

Situation_profession_agreg_CM[1] | 3

	N_menage	Milieu	Region_12	Taille_men-e	Taille_agr-e	Sexe_CM	Age_CM	Age_quin_CM	Lieunaissa-M	Etat_matri-M	Niveau_sco-M	Diplome_ag-M	Situation_-M	DAM
1	1	Urbain	Dakhla-Ou...	2	2 personnes	Masculin	26	25-29	1	Mari□	Primaire	Niveau moyen	Ind□pendant	78027.5
2	2	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	36	35-39	1	Mari□	Secondair...	Niveau moyen	Salari□	47075.4
3	3	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	38	35-39	2	Mari□	Primaire	Sans dipl□me	Ind□pendant	60254
4	4	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	34	30-34	1	Mari□	Sup□rieur	Niveau su...	Salari□	157237
5	5	Urbain	Dakhla-Ou...	4	4 personnes	Masculin	42	40-45	2	Mari□	Sup□rieur	Niveau su...	Salari□	288210
6	6	Urbain	Dakhla-Ou...	5	5 personnes	Masculin	33	30-34	1	Mari□	Secondair...	Niveau moyen	Salari□	137810.6
7	7	Urbain	Dakhla-Ou...	4	4 personnes	Masculin	40	40-45	1	Mari□	Secondair...	Niveau moyen	Salari□	154700.8
8	8	Urbain	Dakhla-Ou...	6	6 personn...	Masculin	50	50-54	2	Mari□	Secondair...	Niveau moyen	Ind□pendant	367526.6
9	9	Urbain	Dakhla-Ou...	8	6 personn...	Masculin	49	45-49	2	Mari□	Secondair...	Niveau moyen	Salari□	57075.5
10	10	Urbain	Dakhla-Ou...	8	6 personn...	Masculin	53	50-54	2	Mari□	Sans niveau	Sans dipl□me	Ind□pendant	165496.8
11	11	Urbain	Dakhla-Ou...	2	2 personnes	Masculin	26	25-29	1	Mari□	Primaire	Niveau moyen	Ind□pendant	80433
12	12	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	36	35-39	1	Mari□	Secondair...	Niveau moyen	Salari□	40024.4
13	13	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	38	35-39	2	Mari□	Primaire	Sans dipl□me	Ind□pendant	62994
14	14	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	34	30-34	1	Mari□	Sup□rieur	Niveau su...	Salari□	158523
15	15	Urbain	Dakhla-Ou...	4	4 personnes	Masculin	42	40-45	2	Mari□	Sup□rieur	Niveau su...	Salari□	282582
16	16	Urbain	Dakhla-Ou...	5	5 personnes	Masculin	33	30-34	1	Mari□	Secondair...	Niveau moyen	Salari□	124083.6
17	17	Urbain	Dakhla-Ou...	4	4 personnes	Masculin	40	40-45	1	Mari□	Secondair...	Niveau moyen	Salari□	152054.8
18	18	Urbain	Dakhla-Ou...	6	6 personn...	Masculin	50	50-54	2	Mari□	Secondair...	Niveau moyen	Ind□pendant	375112.6
19	19	Urbain	Dakhla-Ou...	8	6 personn...	Masculin	49	45-49	2	Mari□	Secondair...	Niveau moyen	Salari□	59947.5
20	20	Urbain	Dakhla-Ou...	8	6 personn...	Masculin	53	50-54	2	Mari□	Sans niveau	Sans dipl□me	Ind□pendant	172050
21	21	Urbain	Dakhla-Ou...	2	2 personnes	F□minin	65	65-69	1	Divorc□	Secondair...	Niveau moyen	Inactif	60618.6
22	22	Urbain	Dakhla-Ou...	2	2 personnes	Masculin	57	55-59	2	Mari□	Sans niveau	Sans dipl□me	Inactif	88721.4
23	23	Urbain	Dakhla-Ou...	3	3 personnes	Masculin	31	30-34	1	Mari□	Secondair...	Niveau su...	Salari□	96986.8
24	24	Urbain	Dakhla-Ou...	4	4 personnes	Masculin	39	35-39	2	Mari□	Secondair...	Niveau moyen	Salari□	153817.4
25	25	Urbain	Dakhla-Ou...	4	4 personnes	Masculin	55	55-59	1	Mari□	Autre niveau	Sans dipl□me	Non d□clar□	135786
26	26	Urbain	Dakhla-Ou...	4	4 personnes	F□minin	46	45-49	2	Veuf	Secondair...	Niveau su...	Salari□	210212.4
27	27	Urbain	Dakhla-Ou...	5	5 personnes	Masculin	46	45-49	1	Mari□	Secondair...	Niveau su...	Salari□	214537.4
28	28	Urbain	Dakhla-Ou...	5	5 personnes	Masculin	47	45-49	1	Mari□	Secondair...	Niveau su...	Salari□	116665.8
29	29	Urbain	Dakhla-Ou...	7	6 personn...	Masculin	67	65-69	1	Mari□	Sans niveau	Sans dipl□me	Salari□	213831
30	30	Urbain	La□youne-...	3	3 personnes	Masculin	39	35-39	1	Mari□	Sup□rieur	Niveau su...	Salari□	177038.2
31	31	Urbain	La□youne-...	3	3 personnes	Masculin	36	35-39	1	Mari□	Sup□rieur	Niveau su...	Salari□	165275.4
32	32	Urbain	La□youne-...	4	4 personnes	Masculin	35	35-39	1	Mari□	Secondair...	Niveau moyen	Salari□	74849.6
33	33	Urbain	La□youne-...	3	3 personnes	Masculin	39	35-39	1	Mari□	Secondair...	Niveau moyen	Ind□pendant	36417

- Les données d'une étude statistique portent sur les individus et peuvent être de plusieurs natures (qualitative, quantitative discret et quantitative continue) ;
- Pour analyser la relation entre deux variables de la base, nous devons faire sortir un autre tableau statistique à partir de cette base ;
- Souvent, ce tableau s'appelle **Tableau de contingence**, surtout pour les variables qualitatives ;

Types de croisements entre caractères relatifs aux même unités

Y X	Quantitatif Discret	Quantitatif Continu	Qualitatif
Quantitatif Discret	Même type		**Exemple 1**
Quantitatif Continu		Même type	
Qualitatif		Exemple 2	Même type

- Exemple 1 : Y : Taille du ménage et X : état matrimoniale du chef du ménage
- Exemple 2 : Y : Revenu du ménage et X : Bénéfice d'un programme social (Oui/Non)

Présentation d'un tableau statistique à double entrées

	Modalités de Y								
Modalités de X		MY_1	...	...	MY_J	...	...	MY_q	Total
	MX_1	n_{11}	...	...	n_{1j}	...	...	n_{1q}	$n_{1.}$
	...								...
	...								...
	MX_i	n_{i1}	...	...	n_{ij}	...	...	n_{iq}	$n_{i.}$
	...								...
	...								...
	MX_q	n_{p1}	...	...	n_{pj}	...	...	n_{pq}	$n_{p.}$
	Total	$n_{.1}$	...	...	$n_{.j}$	...	...	$n_{.q}$	$n_{..}=N$

Lecture du tableau

Que signifie chaque cellule du tableau ?

- n_{11} = card (individus / $X_h = MX_1$ et $Y_h = MY_1$) Pour les individus de 1 à N ;
- $n_{1.}$ = card (individus / $X_h = MX_1$) Pour les individus de 1 à N ;
- n_{i1} = card (individus / $X_h = MX_i$ et $Y_h = MY_1$) Pour les individus de 1 à N ;
- n_{ij} = card (individus / $X_h = MX_i$ et $Y_h = MY_j$) Pour les individus de 1 à N ;
- n_{pq} = card (individus / $X_h = MX_p$ et $Y_h = MY_q$) Pour les individus de 1 à N ;
- $n_{.j}$ = card (individus / $Y_h = MY_j$) Pour les individus de 1 à N ;
- $n_{..} = \sum_{h=1}^{q} n_{.h} = \sum_{h=1}^{p} n_{h.} = N$

Distribution conjointe (X,Y)

	MY_1	...	...	MY_J	...	...	MY_q	Total
MX_1	f_{11}	...	...	f_{1j}	...	...	f_{1q}	$f_{1.}$
...								...
...								...
MX_i	f_{i1}	...	...	f_{ij}	...	...	f_{iq}	$f_{i.}$
...								...
...								...
MX_q	f_{p1}	...	...	f_{pj}	...	...	f_{pq}	$f_{p.}$
Total	$f_{.1}$	...	...	$f_{.j}$	...	...	$f_{.q}$	$f_{..} = 1$

- $f_{ij} = \frac{n_{ij}}{n_{..}}$
- $f_{.j} = \frac{n_{.j}}{n_{..}}$
- $\sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} = \frac{n_{..}}{n_{..}} = 1$

Distribution conditionnelle X/Y et Y/X

- **Distribution conditionnelle X/Y**

	MY_1	...	...	MY_J	...	...	MY_q	Total
MX_i	$f_{1/j}$	...	...	$f_{j/i}$	...	...	$f_{q/i}$	1

1. $f^C_{i/j} = \frac{n_{ij}}{n_{i.}} = \frac{f_{ij}}{f_{i.}}$
2. $\sum_{i=1}^{p} f^C_{i/j} = 1$

Distribution conditionnelle X/Y et Y/x

- **Distribution conditionnelle Y/X**

				MY_J				
MX_1				$f_{1/j}$				
...								
...								
MX_i				$f_{i/j}$				
...								
...								
MX_q				$f_{p/j}$				
Total				1				

1. $f^L_{i/j} = \frac{n_{ij}}{n_{.j}} = \frac{f_{ij}}{f_{.j}}$
2. $\sum_{i=1}^{p} f^L_{i/j} = 1$

Notion d'Indépendance, de Corrélation et de liaison fonctionnelle

- Étude simultanée de deux variables X et Y définies sur une même population P : mettre en évidence une éventuelle liaison (relation, dépendance) entre les variables ;
- Exemples : Etude de la liaison entre
  1. Le revenu du père et le revenu du fils (quantitatives) ;
  2. Le revenu du ménage et le sexe du chef du ménage (quantitative / qualitative) ;
  3. L'appréciation de l'amélioration du niveau de vie et le milieu de résidence (qualitatives).

Définition de l'indépendance d'une fonction à une variable

- La relation entre deux variables peut être :
  1. **Indépendance : La connaissance de X ne permet pas de prévoir Y**
  2. **Liaison fonctionnelle : La connaissance de X détermine Y**
  3. **Corrélation statistique : La connaissance de X donne des idées sur Y**
- **Rôle des variables dans la relation :** dans certains cas, une variable peut en expliquer une autre, dans d'autres cas, les variables jouent des rôles symétriques.
- **Vocabulaire :** Pour des variables qualitatives : **association**. Pour des variables quantitatives : **corrélation**.

Définition de l'indépendance statistique

L'indépendance statistique renferme deux notions :

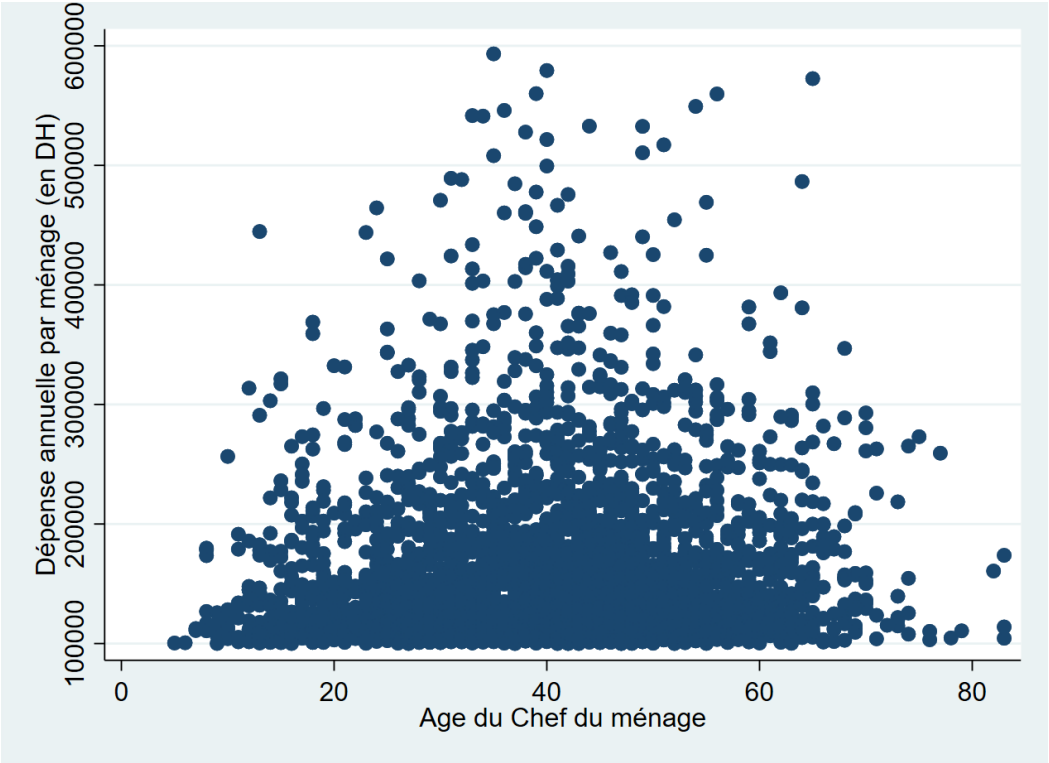
1. L'indépendance d'une fonction à une variable :
   - La valeur de Y est constante Y^* quelque soit la valeur de X ;
   - f(X) = Y^* : la fonction f est indépendante de X ;
2. L'indépendance statistique :
   - Les distributions conditionnelles de X/Y sont identique : $f^C_{i/j} = f^C_{i/j^*}$
   - La conséquence de ceci est qu'elles sont identiques à la distribution marginale de X : $\frac{n_{i1}}{n_{.1}} = \frac{n_{i2}}{n_{.2}} = ... = \frac{n_{iq}}{n_{.q}} = \frac{n_{i.}}{n_{..}}$;
   - $f^C_{i/j} = f_{i.}$;

INSEA

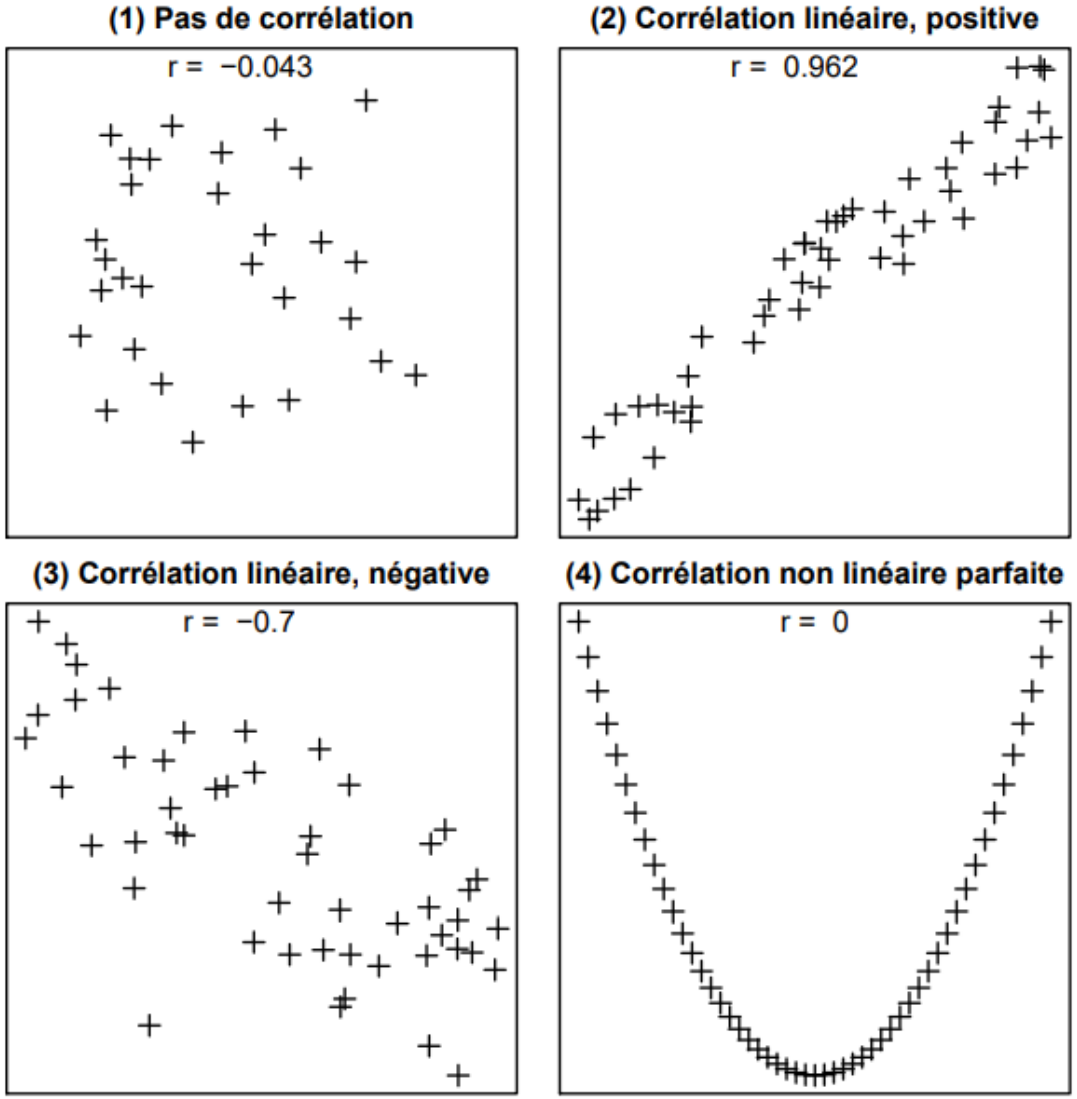
Indépendance et corrélation pour les variables quantitatives

Analyse descriptive de la corrélation : représentation graphique

- Pour avoir une idée sur la relation entre les deux variables, on établi le graphique de **nuage de points ou diagramme de dispersion (scatter-plot)** ;
- Les deux variables sont représentées par chacun des axes.
- Exemple :



Analyse descriptive de la corrélation : représentation graphique



Indépendance et corrélation pour les variables quantitatives

Analyse descriptive de la corrélation : la covariance

- Le terme covariance renvoie à la variance. La variance est une mesure qui quantifie la dispersion moyenne des valeurs prises par cette variable autour de sa moyenne ;
- Deux variables co-varient ensemble lorsqu'un écart à la moyenne d'une variable est accompagné par un écart dans le même sens ou dans le sens opposé de l'autre variable ;
- Autrement dit, deux variables co-varient lorsque la variation d'une des variables autour de sa moyenne semble influencer la manière dont l'autre variable varie autour de sa moyenne ;
- La covariance exprime donc une quantité de variance partagée entre deux variables. En effet, tout comme la variance, plus la valeur de la covariance est élevée, plus les deux variables partagent une portion importante de variance.

INSEA

Indépendance et corrélation pour les variables quantitatives

Analyse descriptive de la corrélation : la covariance

- La covariance se calcule en utilisant la formule suivante :

$$\sigma_{X,Y} = Cov(X, Y) = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{N} = \frac{\sum_{i=1}^{N} X_i . Y_i}{N} - \bar{X}.\bar{Y}$$

- Propriétés de la covariance :
  1. Cov(X,Y) = Cov(Y, X) ;
  2. Cov(X,X) = Var(X) ;
  3. Var(X+Y) = Var(X) + Var(Y) + 2 Cov(X,Y) ;
  4. $\forall\ a, b, c, x_0, y_0 \in R : Cov(aX + x_0, bY + y_0) = a.b.Cov(X, Y)$
     $\Rightarrow Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2.a.b.Cov(X, Y)$
  5. $|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)}$
- Le calcul de la covariance diffère entre les données d'une population et l'échantillon ;
- Les applications statistiques permettent le calcul des matrices de covariances des variables quantitatives ;

INSEA

Indépendance et corrélation pour les variables quantitatives

Analyse descriptive de la corrélation : la covariance

- Exemple : soit le tableau suivant relatif à une étude sur 10 semaines sur les retombés de la publicité d'une société sur ses ventes :

Semaine	Nombre de spots publicitaires	Ventes en unités
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

- Calculer la covariance entre le nombre de sports publicitaires (X) et les ventes (Y)

Indépendance et corrélation pour les variables quantitatives

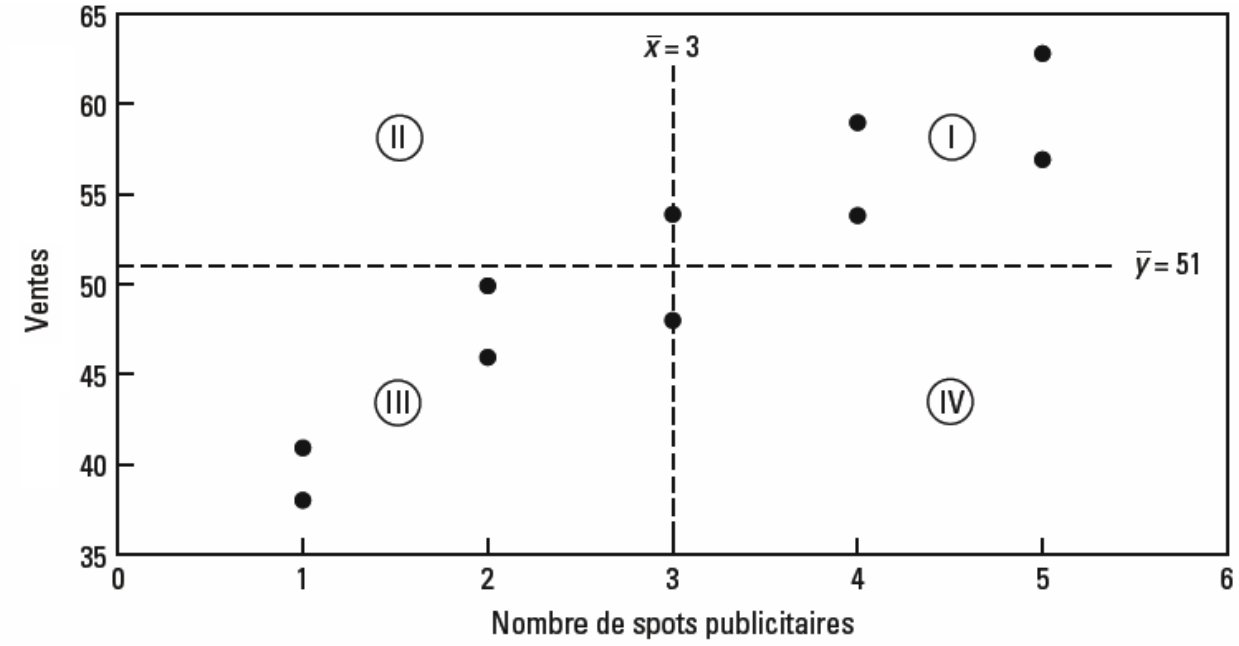
Analyse descriptive de la corrélation : la covariance

Semaine	**Spots (X)**	**Ventes (Y)**	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X}).(Y_i - \bar{Y})$
1	2	50	-1	-1	1
2	5	57	2	6	12
3	1	41	-2	-10	20
4	3	54	0	3	0
5	4	54	1	3	3
6	1	38	-2	-13	26
7	5	63	2	12	24
8	3	48	0	-3	0
9	4	59	1	8	8
10	2	46	-1	-5	5
Total	**30**	**510**			**99**

- $\bar{X} = 3$ et $\bar{X} = 51$
- Cov(X,Y) $= \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{N} = \frac{99}{10} = 9.9$

Indépendance et corrélation pour les variables quantitatives

Interprétation de la covariance



- Si la valeur de $\sigma_{X,Y}$ est positive, les points qui ont la plus grande influence sur $\sigma_{X,Y}$ se trouvent dans les cadrans I et III ;
- Une valeur positive de $\sigma_{X,Y}$ révèle une relation linéaire positive entre X et Y ; c'est-à-dire, lorsque la valeur de X augmente, la valeur de Y augmente ;
- Si les points sont répartis de façon uniforme entre les quatre cadrans, la valeur de $\sigma_{X,Y}$ sera proche de zéro, indiquant l'absence d'une relation linéaire entre X et Y.

La corrélation

- Malgré que la covariance donne une idée sur la liaison entre deux variables quantitatives, elle n'est pas facilement lisible ;
- Pour une meilleure lecture de cette liaison, on fait appel à une autre mesure qui se vase sur la covariance : c'est le coefficient de **corrélation** ;
- Selon les données que nous utilisons, il y a plusieurs mesure de la corrélation, mais la plus célèbre est le coefficient de corrélation de **Pearson** ;
- Mais pour pouvoir utiliser ce coefficient, il faut que les données vérifient 4 hypothèses à savoir : la continuité des données, la linéarité de la liaison entre les deux variables, la normalité des données et l'absence des observations abérantes ;
- **Une corrélation positive entre deux variables ne signifie pas automatiquement une relation de cause à effet. C'est à dire que si deux variables sont positivement liées, cela ne veut pas dire qu'une variable détermine l'autre ;**
- Si les hypothèses ne sont pas vérifiées, on fait appel à d'autres mesures comme le coefficient de **Spearman** ou celui de **Kendall** ;

Le coefficient de corrélation de Pearson

- Par la suite, nous allons assumer la vérification des hypothèses pour utiliser le coefficient de Pearson ;
- Le coefficient se calcule comme suit :

$$r_{X,Y} = \rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X.\sigma_Y} = \frac{Cov_{X,Y}}{Ecart-type(X).Ecart-type(Y)}$$

$$\rho_{X,Y} = \frac{\sum_{i=1}^{N}(X_i-\bar{X}).(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i-\bar{X})^2.\sum_{i=1}^{N}(Y_i-\bar{Y})^2}}$$

$$\rho_{X,Y} = \frac{N.\sum_{i=1}^{N}X_i.Y_i-\sum_{i=1}^{N}X_i\sum_{i=1}^{N}Y_i}{\sqrt{N.\sum_{i=1}^{N}X_i^2-\left(\sum_{i=1}^{N}X_i\right)^2}.\sqrt{N.\sum_{i=1}^{N}Y_i^2-\left(\sum_{i=1}^{N}Y_i\right)^2}}$$

- Au delà du calcul de coefficient de corrélation, il y a des tests statistiques pour vérifier si le coefficient calculé est statistiquement différent de 0 ou non ;

INSEA

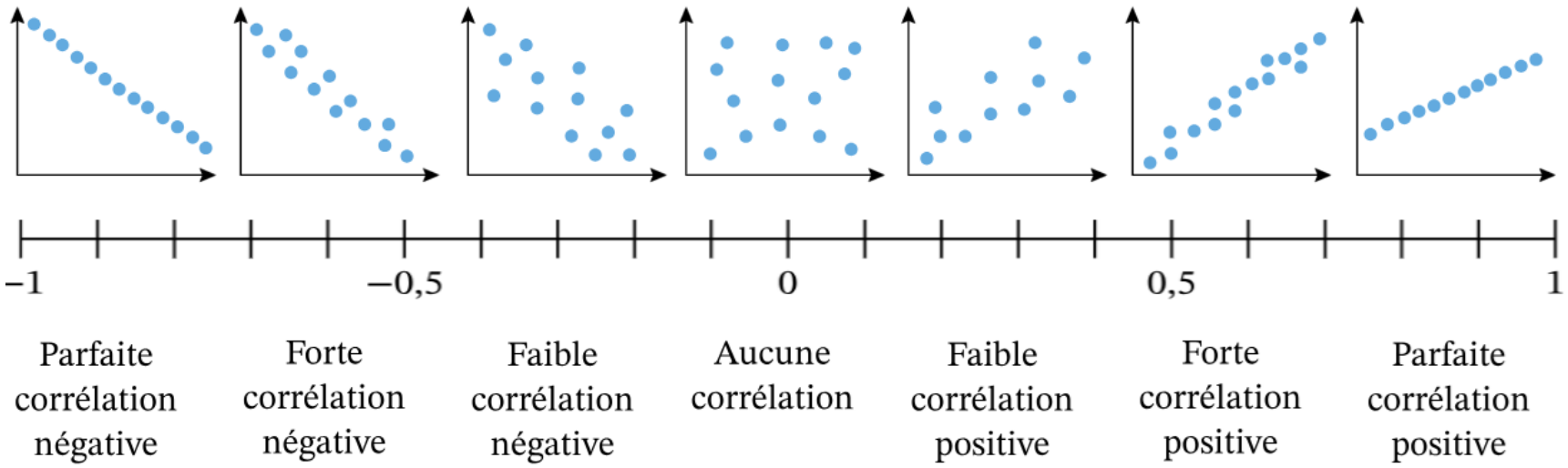
Lecture du coefficient de corrélation de Pearson

- Le coefficient se situe toujours entre -1 et 1 ;
- Si le coefficient est proche de -1 : Il y a une forte liaison inverse entre X et Y. C'est à dire que si X augmente, Y tend à diminuer ;
- Si le coefficient est proche de 1 : Les deux variables X et Y augmentent ou diminuent ensemble. C'est à dire que si X augmente, Y tend à augmenter aussi ;
- Si le coefficient est proche de 0 : Aucune liaison apparente entre les variables. Dans ce cas, on peut visualiser les données pour voir la nature de la liaison entre X et Y et faire les transformations nécessaire de données pour plus de visibilité ;

INSEA

Lecture du coefficient de corrélation de Pearson

- La figure suivante présente la méthode de lecture du coefficient de corrélation.



La coefficient de corrélation

- Exemple 1 : Reprenons l'exemple précédent sur le nombre des spots publicitaires et les ventes d'une entreprise.

Semaine	Nombre de spots publicitaires	Ventes en unités
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

- Calculer le coefficient de corrélation.

Indépendance et corrélation pour les variables quantitatives

Analyse descriptive de la corrélation : la covariance

X	**Y**	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X}).(Y_i - \bar{Y})$
2	50	-1	-1	1	1	1
5	57	2	6	4	36	12
1	41	-2	-10	4	100	20
3	54	0	3	0	9	0
4	54	1	3	1	9	3
1	38	-2	-13	4	169	26
5	63	2	12	4	144	24
3	48	0	-3	0	9	0
4	59	1	8	1	64	8
2	46	-1	-5	1	25	5
30	**510**			**20**	**566**	**99**

- $\bar{X} = 3$, $\bar{X} = 51$, $\sigma_X = 1,49$ et $\sigma_Y = 7,93$
- $Cov(X, Y) = 9.9$
- $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X.\sigma_Y} = \frac{9,9}{1,49.9,93} = 0,84$ **: Une forte liaison positive.**

Liaison entre deux variables qualitatives

- Dans le cas où on a deux variables qualitatives, on étudie la nature de la liaison entre les deux variables à travers :
  1. Coefficient de contingence ;
  2. Coefficient de Pearson

Définition de l'indépendance statistique

L'indépendance statistique renferme deux notions :

1. L'indépendance d'une fonction à une variable :
   - La valeur de Y est constante Y^* quelque soit la valeur de X ;
   - f(X) = Y^* : la fonction f est indépendante de X ;
2. L'indépendance statistique :
   - Les distributions conditionnelles de X/Y sont identique : $f^C_{i/j} = f^C_{i/j^*}$
   - La conséquence de ceci est qu'elles sont identiques à la distribution marginale de X : $\frac{n_{i1}}{n_{.1}} = \frac{n_{i2}}{n_{.2}} = ... = \frac{n_{iq}}{n_{.q}} = \frac{n_{i.}}{n_{..}}$;
   - $f^C_{i/j} = f_{i.}$;

Mesure d'indépendance : Coefficient de contingence

- La relation d'indépendance est la suivante : $\frac{f_{ij}}{f_{.j}} = f_{i.}$
  $\Rightarrow f_{ij} = f_{.j}.f_{i.}$
  $\Rightarrow \delta_{ij} = f_{ij} - f_{.j}.f_{i.} = 0$
- On définit le coefficient de contingence C comme suit :
  $\Phi^2_{ij} = \frac{\delta^2_{ij}}{f_{i.}.f_{j.}}$
  $\Rightarrow \Phi^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \Phi^2_{ij}$
- Alors le coefficient C est égale :
  $C = \frac{\Phi^2}{min(p,q)-1}$; $\mathbf{C \geq 0}$

Mesure d'indépendance : Coefficient de Pearson

- Un autre coefficient est préconisé pour mesurer l'indépendance, c'est le coefficient de **Pearson :**
- $P = \sqrt{\frac{\Phi^2}{1+\Phi^2}}$
- $\mathbf{P \geq 0}$

INSEA

Calcul pratique des coefficients

- Pour calculer le coefficient de contingence à partir des effectifs du tableau statistique, on utilise la formule simplifiée :

$$\Phi^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{n_{ij}^2}{n_{i.} . n_{.j}} - 1$$

- Exemple : soit le tableau de contingence suivant entre deux variables X et Y :

	X_1	X_2	X_3	X_4	X_5	**Total**
Y_1	15	35	12	24	19	**105**
Y_2	21	20	14	11	17	**83**
Y_3	18	17	14	29	34	**112**
Y_4	21	13	24	17	12	**87**
Y_5	25	24	18	26	30	**123**
Total	**100**	**109**	**82**	**107**	**112**	**510**

- Calculer les indices Φ^2, C et P

Calcul pratique des coefficients

	X_1	X_2	X_3	X_4	X_5	**Total**
Y_1	15	35	12	24	19	**105**
Y_2	21	20	14	11	17	**83**
Y_3	18	17	14	29	34	**112**
Y_4	21	13	24	17	12	**87**
Y_5	25	24	18	26	30	**123**
Total	**100**	**109**	**82**	**107**	**112**	**510**

1. $\Phi^2 = \frac{15^2}{100.105} + \frac{35^2}{109.105} + \ldots + \frac{19^2}{112.105} + \frac{21^2}{100.83} + \ldots + \frac{17^2}{112.83} + \ldots + \frac{30^2}{100.105} - 1 = 0,07$
2. $C = \frac{\Phi^2}{min(p,q)-1} = \frac{0.07}{5-1} = 0.175$
3. $P = \sqrt{\frac{\Phi^2}{1+\Phi^2}} = \sqrt{\frac{0.07}{1+0.07}} = 0.25$

INSEA

Mesures de Corrélation : Liaisons fonctionnelles

- Quand l'indépendance entre X et Y n'est pas vérifié, on peut rencontrer les cas extrêmes de liaisons fonctionnelles
  1. De Y en X (dans ce cas $p > q$) ;
  2. De X en Y (dans ce cas $p < q$) ;
  3. Réciproque (dans ce cas $p = q$).

Relation entre une variable qualitative et une variable quantitative

- Pour le cas où on une variable qualitative et une quantitative, les deux variables ne jouent pas un rôle symétrique ;
- On veut étudier l'influence d'une variable qualitative (par exemple le sexe) sur une variable quantitative (par exemple le salaire) ;
- Dans ce cas, on peut utiliser les mêmes coefficient de contingence C, de Pearson P comme le cas de deux variables qualitatives ;
- Mais, **la meilleure technique de faire une analyse ANOVA : analyse de la variance à un facteur.**

.

Questions ?